Towards Mathematical Sciences
Vol. 2, No. 1, Spring and Summer 2022, pp. 110–121
DOI.10.22067/tmsj.2022.42824
Expository Article

به سوی علوم ریاضی

https://tmsj.um.ac.ir

REDUCING SEARCH SPACE IN OPTIMIZATION PROBLEMS WITH DATA MINING TECHNIQUES

MAHTAB HADDADPOUR[1], MOHAMMAD ALINEJADMOFRAD[1*], AND MOHAMMAD DEHGHAN NAYYERI[1]

[1]Department of Basic sciences, University of Bojnord, Bojnord, Iran
m.haddadpour@stu.ub.ac.ir
m.alinejad@ub.ac.ir
m.dehghan@ub.ac.ir

Abstract. Finding the optimal global solution in optimization problems is such an important issue that various related approaches have been proposed so far. An effective attempt before solving such problems is to reduce the search space in such a way that the search is concentrated in a smaller subspace and therefore the probability of finding the optimal global solution increases. In this article, three methods of clustering, classification and association in data mining are used to reduce the search space in a nonlinear optimization problem. After that, using the Genetic Algorithm, the problem is solved on the entire initial feasible space and the reduced spaces resulting from three data mining methods. The results show that by combining data mining methods and Genetic Algorithm, more accurate approximations for the global optimal solution of the problem can be obtained.

2020 Mathematics Subject Classification. 90C26, 90C59
Keywords. Optimal global solution, data mining, reduced search space, genetic algorithm
Date: Rceived 16-2-2022 Revised 21-8-2022 Accepted 4-9-2022 Available Online 9-9-2022
*Corresponding author.

دوره ۲، شماره ۱، بهار و تابستان ۱۴۰۱، صص. ۱۲۱–۱۱۰
DOI:10.22067/tmsj.2022.42824
مقاله علمی–ترویجی

https://tmsj.um.ac.ir

کاهش فضای جستجوی مسائل بهینه‌سازی با روش‌های داده‌کاوی

مهتاب حدادپور[۱]، محمّد علی‌نژادمفرد[۱*]، و محمّد دهقان‌نیّری[۱]

[۱] گروه ریاضی، دانشکده علوم پایه، دانشگاه بجنورد
m.haddadpour@stu.ub.ac.ir
m.alinejad@ub.ac.ir
m.dehghan@ub.ac.ir

چکیده. یافتن جواب بهینه سراسری در مسائل بهینه‌سازی، تا اندازه‌ای اهمیت دارد که تاکنون رویکردهای متنوعی برای آن ارائه شده است. یک اقدام موثر قبل از حل این دست از مسائل، کاهش‌دادن (کوچک کردن) فضای جستجو است به نحوی که جستجو در یک زیرفضای کوچکتر متمرکز گردد و احتمال یافتن جواب بهینه سراسری افزایش یابد. در این مقاله از سه روش خوشه‌بندی، طبقه‌بندی و انجمنی در داده‌کاوی برای کاهش فضای جستجو در یک مسئله بهینه‌سازی غیرخطی استفاده می‌شود. پس از آن به کمک الگوریتم ژنتیک، مسئله روی کل فضای شدنی اولیه و فضاهای کاهش‌یافته حاصل از سه روش داده‌کاوی حل می‌شود. نتایج نشان می‌دهند که می‌توان با ترکیب روش‌های داده‌کاوی و الگوریتم ژنتیک، تقریب‌های دقیق‌تری برای جواب بهینه سراسری مسئله به دست آورد.

۰. مقدمه

در عصر حاضر مفهوم بهینه‌سازی به‌عنوان یک اصل اساسی در تجزیه و تحلیل بسیاری از مسائل پیچیده به‌طورکامل پذیرفته شده است. با به‌کارگرفتن نظریه بهینه‌سازی، می‌توان مسائلی را بررسی کرد که شامل تعیین مقادیری برای چند متغیر مرتبط به هم و تاثیرگذار روی هدفی یگانه است. این هدف، برای اندازه‌گیری کمی و

2020 Mathematics Subject Classification. 90C26, 90C59.

واژگان کلیدی. جواب بهینه سراسری، داده‌کاوی، فضای جستجوی کاهش‌یافته، الگوریتم ژنتیک.
تاریخ: دریافت ۱۴۰۱/۲/۲۶ ویرایش ۱۴۰۱/۵/۳۰ پذیرش ۱۴۰۱/۶/۱۳ انتشار برخط ۱۴۰۱/۶/۱۸

*نویسنده مسئول

نحوه ارجاع به این مقاله: م. حدادپور، م. علی‌نژادمفرد، م. دهقان نیری، کاهش فضای جستجوی مسائل بهینه‌سازی با روش‌های داده‌کاوی، به سوی علوم ریاضی، ۲ (۱۴۰۱)، شماره ۱، ۱۲۱–۱۱۰.

کیفی تصمیم ساخته می‌شود و با توجه به قیودی که ممکن است انتخاب مقادیر متغیرهای تصمیم‌گیری را محدود کنند، کمینه یا بیشینه (برحسب فرمول‌بندی تصمیم) می‌گردد [۴].

بهینه‌سازی سراسری[۱] به‌دنبال یافتن جواب بهینه سراسری در فضای شدنی است. سه رویکرد اصلی برای حل این مسئله ارائه شده است؛ این رویکردها شامل استفاده از الگوریتم‌های متناهی[۲] ، روش‌های تکراری همگرا[۳] [۴، ۵] و روش‌های مکاشفه‌ای[۴] است [۹، ۱۲]. روش‌های مکاشفه‌ای در بهینه‌سازی برای حل سریع‌تر مسئله، زمانی‌که روش‌های کلاسیک مبتنی بر دو رویکرد دیگر خیلی کُند هستند، و یا برای محاسبه یک جواب تقریبی (زیربهین)، وقتی‌که روش‌های کلاسیک قادر به محاسبه جواب دقیق نیستند، به‌کار می‌روند؛ و دراین‌صورت، به طورمعمول با صرف‌نظر از بهینگی (بهینه سراسری بودن)، کامل‌بودن (یافتن همه بهینه‌های محلی)، دقت (دقت محاسباتی) و درستی (اعتبار جواب) در برابر سریعتر یافتن جواب به هدف خود نائل می‌شوند [۹]. روش‌های مکاشفه‌ای با جستجو در فضای جستجو جواب مسئله را می‌یابند، و هرچه فضای جستجو کوچکتر باشد سرعت یافتن جواب مسئله و در بیشتر موارد، دقت و درستی جواب یافت‌شده نیز بیشتر می‌شود. با کاهش فضای جستجو[۵] در مسئله بهینه‌سازی، جستجو در یک ناحیه کوچکتر متمرکز می‌گردد و در نتیجه، احتمال یافتن جواب بهینه سراسری نیز افزایش می‌یابد. الگوریتم ژنتیک یکی از این دست از الگوریتم‌هاست که از طبیعت و تئوری انتخاب داروین الگو گرفته است و تاکنون برای حل بسیاری از مسائل بهینه‌سازی با موفقیت استفاده شده است. در این الگوریتم ابتدا به صورت تصادفی یک دسته جواب روی کل فضای شدنی تولید شده و سپس با استفاده از عملگرهای الگوریتم ژنتیک مانند تقاطع[۶] و جهش[۷]، رفته‌رفته جواب‌های بهتری حاصل می‌شود.

داده‌کاوی[۸] با تجزیه و تحلیل یک مجموعه بسیار بزرگ از داده‌ها، قوانین ارزشمند یا الگوهای پنهان در داده را استخراج می‌کند. این استخراج با روش‌هایی همچون طبقه‌بندی[۹]، انجمنی[۱۰] و خوشه‌بندی[۱۱] انجام می‌شود. طبقه‌بندی شیوه‌ای برای تجزیه و تحلیل داده‌ها است که الگویی برای توصیف طبقات و کلاس‌های مهم موجود در داده‌ها ایجاد می‌کند. این قبیل الگوها برچسب کلاس‌ها را پیش‌بینی می‌کنند. قوانین انجمنی، تجزیه و تحلیل الگوهای تکراری به‌منظور کشف قوانین در مجموعه داده‌ها است. خوشه‌بندی فرآیندی برای گروه‌بندی مجموعه‌ای از داده‌ها در چند خوشه (گروه) است که اعضای هر خوشه شبیه به‌هم هستند و با اعضای دیگر خوشه‌ها تفاوت دارند [۲، ۳، ۱].

---

[۱] global optimization

[۲] finitely terminating algorithms

[۳] convergent iterative methods

[۴] heuristics methods

[۵] reducing search space

[۶] crossover

[۷] mutation

[۸] data mining

[۹] classification

[۱۰] association

[۱۱] clustering

در پژوهش [۷]، یک الگوریتم بهینه‌سازی سراسری ترکیبی[۱۲] توسعه یافته که در آن از روش‌های داده‌کاوی برای کاهش فضای جستجو استفاده شده است. رویکرد اتخاذ شده در [۷] برای کاهش فضای جستجو در روش‌های طبقه‌بندی و انجمنی عموما به اجتماع زیرفضاهایی از فضای جواب منجر می‌شود که ممکن است در حالت کلی، جواب بهینه سراسری را شامل نشود. در این تحقیق از رویکردی محافظه‌کارانه برای کاهش فضای جستجو استفاده می‌شود تا تنها یک فضای کاهش‌یافته از هر یک از روش‌های داده‌کاوی حاصل شود و به این ترتیب احتمال وجود بهینه سراسری در فضای کاهش‌یافته افزایش یابد. در ادامه نخست در بخش دو، مفاهیم اولیه مورد نیاز در بهینه‌سازی سراسری بیان می‌شود، سپس در بخش سه نحوه آماده‌سازی داده‌های ورودی برای نرم‌افزار وکا[۱۳] شرح داده می‌شود. در بخش چهار، یک مسئله بهینه‌سازی غیرخطی معرفی و کاهش فضای جستجو با روش‌های متنوع داده‌کاوی روی آن بررسی می‌شود. در بخش پنج برای حل مسئله، الگوریتم ژنتیک با روش‌های داده‌کاوی ترکیب می‌شود و سرانجام در بخش پایانی نتیجه این اقدام موثر، جمع‌بندی و بیان می‌شود. نتایج نشان می‌دهند که اولا دو فضای کاهش‌یافته دقیقا شامل جواب بهینه سراسری مسئله است و دوما با کوچکتر شدن فضاهای جستجو حاصل از روش‌های داده‌کاوی و استفاده از الگوریتم ژنتیک، تقریب‌های بهتری برای جواب بهینه سراسری مسئله به دست می‌آید.

۲. مفاهیم اولیه

مسئله کلی بهینه‌سازی را می‌توان به شکل زیر بیان کرد:

$$
\begin{aligned}
\min \quad & f(X) \\
\text{subject to} \quad & g_i(X) \le 0, \quad i = 1,2,3,...,p, \\
& h_j(X) = 0, \quad j = 1,2,3,...,q, \\
& X \in S,
\end{aligned}
\tag{۱.۲}
$$

که در آن $X = (x_1, x_2, x_3, ..., x_n)$ یک بردار n بعدی از متغیرهای مجهول، g_i ,$i = 1,2,3,...,p$ و h_j ,$j = 1,2,3,...,q$ و f، توابع حقیقی از متغیرهای $x_1, x_2, ..., x_n$ هستند. مجموعه S زیرمجموعه‌ای از فضای n بعدی $\mathbb{R}^n$ است. تابع f تابع هدف مسئله است و معادله‌ها، نامعادله‌ها و محدودیت‌های مجموعه‌ای، قیود مسئله هستند [۴]. ناحیه‌ای که به قیود مسئله محدود می‌شود فضای شدنی نام دارد. این ناحیه شامل تمام جواب‌های ممکن برای مسئله است که در همه قیود صدق می‌کنند. برای به دست آوردن جواب بهینه مسئله، کافی است که از فضای شدنی، آن جوابی که تابع هدف را کمینه (یا بیشینه) می‌کند، یافت.

[۱۲] hybrid global optimization algorithm

[۱۳] Weka

تعریف ۱.۲. نقطه $X^* \in S$ را کمینه سراسری[۱۴] تابع f روی S گویند هرگاه برای هر $X \in S$ داشته باشیم $f(X^*) \leq f(X)$ و آن را کمینه محلی[۱۵] تابع f روی S گویند هرگاه $\epsilon > 0$ موجود باشد به‌طوری‌که برای هر $X \in S \cap N_\epsilon(X^*)$داشته باشیم $f(X^*) \leq f(X)$ که در آن $N_\epsilon(X^*)$ همسایگی‌ای از X^* به شعاع ϵ است.

تعریف ۲.۲. نقطه X^* از فضای شدنی مسئله (۱.۲) را جواب بهینه سراسری[۱۶] آن مسئله گویند هرگاه کمینه سراسری تابع f روی فضای شدنی مسئله (۱.۲) باشد؛ و آن را جواب بهینه محلی مسئله (۱.۲) گویند هرگاه کمینه محلی تابع f روی فضای شدنی آن مسئله باشد.

هرگاه در مسئله (۱.۲)، تابع هدف و تمامی قیود، توابعی خطی باشند، مسئله را خطی و در غیر این صورت، غیرخطی می‌نامند. برای مسائل خطی، جواب بهینه محلی همان جواب بهینه سراسری است اما برای مسائل غیرخطی چنین نیست و نیز بسیاری از الگوریتم‌ها همواره قادر نیستند که جواب بهینه سراسری را بیابند. در بیشتر مسائل، تشخیص جواب بهینه سراسری سخت و تعیین آن حتی سخت‌تر است [۸].

۳. آماده‌سازی داده‌ها

برای کاهش فضای جستجو با روش‌های داده‌کاوی، نخست جواب‌هایی به‌صورت تصادفی با توزیع یکنواخت تولید می‌شوند. این جواب‌ها در واقع نقاطی متعلق به تمام فضای شدنی هستند. سپس مقادیر تابع هدف برای هر یک از این نقاط محاسبه می‌شود. آنگاه این مجموعه از نقاط و مقادیر تابع هدف متناظر (به عنوان مجموعه داده‌های ورودی) در نرم‌افزار داده‌کاوی وکا فراخوانی می‌شود تا روش‌های داده‌کاوی روی آن پیاده‌سازی گردد و قوانین یا الگوها به دست آید. نرم‌افزار وکا یکی از معروف‌ترین و پرکاربردترین نرم‌افزارهای حوزه داده‌کاوی است [۱۱]. این نرم‌افزار رایگان است و توسط ویتن[۱۷] و فرنک[۱۸] در دانشگاه ویکاتو[۱۹] از کشور زلاندنو[۲۰] توسعه داده شده است.

فراخوانی داده‌های ورودی در نرم‌افزار برای روش خوشه‌بندی بدون تغییر شکل آنها است. بنابراین در این روش، مقادیر اصلی تابع هدف مورد استفاده قرار می‌گیرد تا فاصله بین این مقادیر محاسبه شده و خوشه‌ها تشکیل شوند. خروجی خوشه‌بندی، تعدادی خوشه به‌همراه میانگین هر خوشه و تعداد اعضای متعلق به آن خوشه است. فراخوانی داده‌ها در نرم‌افزار برای روش‌های طبقه‌بندی و انجمنی نیاز به تغییر شکل آنها دارد. می‌توان از خوشه‌های به‌دست‌آمده از خروجی خوشه‌بندی برای آماده‌سازی داده‌های ورودی روش‌های طبقه‌بندی و انجمنی استفاده کرد. به این ترتیب که بر اساس مقادیر تابع هدف و خروجی خوشه‌بندی، جواب‌های تصادفی

---
[۱۴] global minimum
[۱۵] local minimum
[۱۶] global optimal solution
[۱۷] Witten
[۱۸] Frank
[۱۹] Waikato University
[۲۰] New Zealand

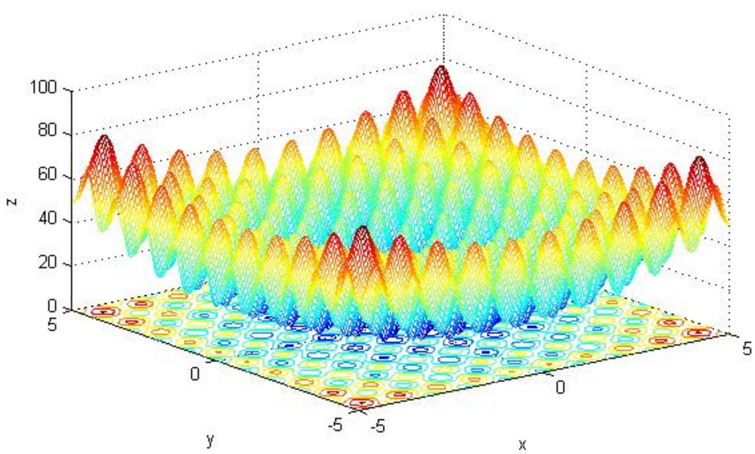
تولید شده به کلاس‌های مختلف (با فواصل یکسان یا نابرابر) تقسیم می‌شوند. بهترین کلاس با نماد A و کلاس‌های دیگر با B، C، D و غیره مطابق با ترتیب مقادیر تابع هدف آنها نمایش داده می‌شود. اکنون این نمادها برای جایگزینی مقادیر تابع هدف در فایل ورودی برای این دو روش استفاده می‌شود. در بخش بعد همین فرآیند را برای آماده‌سازی داده‌ها در پیش می‌گیریم.

۴. کاهش فضای جستجو

برای نمایش چگونگی کاهش فضای جستجو با استفاده از روش‌های داده‌کاوی، مسئله بهینه‌سازی غیرخطی زیر را در نظر بگیرید. مسئله عبارت است از

$$\begin{aligned} \min \quad & f(X) = 20 + x_1^2 + x_2^2 - 10[\cos(2\pi x_1) + \cos(2\pi x_2)] \\ \text{subject to} \quad & -5 \le x_1 \le 5, \\ & -5 \le x_2 \le 5. \end{aligned} \tag{۱.۴}$$

تابع هدف، تابع راستریگین[۲۱] است [۶] و شکل ۱ نمودار آن را نشان می‌دهد. این تابع تنها یک جواب کمینه سراسری در $X^* = (۰،۰)$ و تعداد زیادی کمینه محلی در فضای شدنی دارد که برخی از آنها بسیار نزدیک به جواب بهینه سراسری هستند. این مسئله، یک مسئله بسیار نویزی است و یافتن جواب بهینه سراسری این مسئله با استفاده از روش‌های کلاسیک یا حتی روش‌های مکاشفه‌ای، دشوار است.

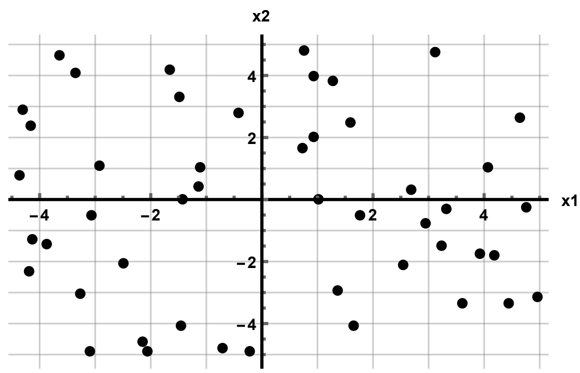


شکل ۱: تابع راستریگین.

برای شروع کار، ابتدا 48 نقطه به صورت تصادفی و با توزیع یکنواخت در کل فضای شدنی تولید می‌شوند. توزیع این نقاط روی صفحه در شکل ۲ نشان داده شده است. حال بر اساس «توزیع» مقادیر تابع هدف این

[۲۱] Rastrigin function

نقاط، چهار کلاس با نمادهای A، B، C و D به این نقاط اختصاص داده می‌شوند. کلاس A شامل چهار عضو با میانگین مقادیر تابع هدف 1.24635 است که بهترین کلاس محسوب می‌شود زیرا میانگین مقادیر تابع هدف متناظر با نقاط آن نسبت به سایر کلاس‌ها، کمترین است. کلاس دوم، B، شامل 25 عضو با میانگین مقادیر تابع هدف 31.0151 است. کلاس سوم، C، دارای 15 عضو با میانگین مقادیر تابع هدف 46.6874 است و بالاخره کلاس D متشکل از چهار عضو با میانگین مقادیر تابع هدف 60.9127 است. در اینجا کلاس‌ها را براساس خوشه‌های خروجی روش خوشه‌بندی در نظر گرفته‌ایم؛ اگرچه می‌توان آنها را به هر طریق دیگری نیز در نظر گرفت که دراین‌صورت کلاس‌بندی‌های گوناگون به نتایج متفاوتی در کاهش فضای جستجو منجر خواهد شد.



شکل ۲: نقاط تصادفی تولید شده در $[-5, 5] \times [-5, 5]$.

خروجی روش‌های خوشه‌بندی، طبقه‌بندی و انجمنی حاصل از نرم‌افزار وکا به ترتیب در جداول ۱، ۲ و ۳ نشان داده شده است. مطابق جدول ۱، چهار خوشه یافت شده است و از آنجا که خوشه ۱، دارای کمترین میانگین مقادیر تابع هدف است، چهار عضو خوشه ۱ برای تعیین مرز این خوشه استفاده می‌شود. چون بزرگ‌ترین مقدار x_1 برای این چهار عضو در خوشه ۱ برابر 1.01996 و کوچکترین مقدار x_1 برای این چهار عضو برابر -2.93562 است از این رو کران‌های x_1 را به صورت $-2.93562 \leq x_1 \leq 1.01996$ در نظر می‌گیریم. به طور مشابه برای x_2 داریم: $0.025382 \leq x_2 \leq 2.00188$. معمولاً، تنها یک فضای کاهش‌یافته توسط روش خوشه‌بندی به دست می‌آید. مساحت زیرفضای حاصل از این روش نزدیک به 8 درصد از مساحت کل فضای جستجو را در بر می‌گیرد. در اینجا خوشه‌بندی به روش اولین–دورترین[۲۲] پیمایش روی گرافی که رئوس آن مقادیر تابع هدف و وزن یال‌ها اختلاف مقادیر رئوس با هم است، انجام شده است [۱۰].

نمایش خروجی طبقه‌بندی معمولاً یک درخت تصمیم شامل تعدادی گره و شاخه است. درخت می‌تواند از ریشه تا یک برگ پیموده شود تا کلاس یک جواب تعیین گردد. در هر گره، برخی از قوانین برای تصمیم‌گیری

[۲۲] Farthest–First

جدول ۱: خروجی خوشه‌بندی

خوشه	میانگین مقادیر تابع هدف	تعداد اعضای خوشه (درصد از کل)
0	60.9127	4 (8%)
1	1.24635	4 (8%)
2	31.0151	25 (52%)
3	46.6874	15 (31%)

در مورد ورود به شاخه‌های مختلف فراهم می‌شود. با استفاده از این قوانین، آن جواب می‌تواند به یک کلاس تخصیص داده شود. جدول ۲ درخت تصمیم متناظر را نشان می‌دهد.

در هر برگ، نماد یک کلاس به همراه یک کسر درون پرانتز نمایش داده شده است؛ صورت کسر، تعداد جواب‌هایی است که قوانین منتهی به آن برگ را برآورده می‌کنند و مخرج کسر تعداد جواب‌هایی است که به آن کلاس تعلق ندارند. از آنجا که کلاس A همان بهترین کلاس است، می‌توان درخت را به سادگی از چپ به راست پیمود و با گذر از گره‌ها و شاخه‌ها به برگ‌هایی با برچسب A رسید. قوانین نمایش داده شده در هر گرهی که به کلاس A منتهی می‌شود، فضای جستجوی کاهش‌یافته را به صورت اجتماع سه فضای $-4.16 \le x_1 < -2.04$ و $0 \le x_2 < 2.25$ همچنین $-1.13 \le x_1 < -0.2$ و $0 \le x_2 < 2.25$ و سرانجام $0.83 \le x_1 < 1.63$ و $0 \le x_2 < 2.25$ نشان می‌دهد. با رویکردی محافظه‌کارانه، چون بزرگ‌ترین مقدار x_1 برای این سه زیرفضا حداکثر 1.63 و کوچکترین مقدار x_1 برای این سه زیرفضا برابر -4.16 است از این رو کران‌های x_1 را به صورت $-4.16 \le x_1 < 1.63$ در نظر می‌گیریم. به طور مشابه برای x_2 داریم: $0 \le x_2 < 2.25$. حال این فضای به دست آمده تنها حدود 12 درصد از کل فضای جستجو است.

نمایش خروجی انجمنی بر خلاف طبقه‌بندی، شامل تعداد بسیاری از قوانین به جای درخت تصمیم است. استفاده از این قوانین باید براساس تعداد جواب‌های صادق در قانون و دقت آن باشد. جدول ۳ خروجی روش انجمنی در نرم‌افزار وکا را نشان می‌دهد. در هر قانون، مثلا عبارت $F = A$ به این معنی است که آن قانون به کلاس A منجر می‌شود و عدد بعد از A تعداد جواب‌هایی است که در آن قانون صدق می‌کند. در انتهای هر قانون، عدد داخل پرانتز، کسری از اطمینان را نمایش می‌دهد. صورت این کسر تعداد جواب‌هایی است که متعلق به کلاس A است و مخرج این کسر تعداد جواب‌هایی است که در آن قانون صدق می‌کند؛ بنابراین، 1 نشان‌دهنده اطمینان صددرصدی به آن قانون است. در اینجا برای این مسئله، از قوانین شماره‌ی 16، 32، 36 و 42 می‌توان برای شناسایی جواب‌های متعلق به کلاس A استفاده کرد. یعنی، فضای کاهش‌یافته به ترتیب توسط اجتماع فضاهای $-3 \le x_1 \le -2$ و $1 \le x_2 \le 2$ همچنین $1 \le x_1 \le 2$ و $0 \le x_2 \le 1$ افزون بر این $0 \le x_1 \le 1$ و $2 \le x_2 \le 3$ و سرانجام $-2 \le x_1 \le -1$ و $1 < x_2 \le 2$، مشخص می‌شود. با رویکردی محافظه‌کارانه، چون بزرگ‌ترین مقدار x_1 برای این چهار زیرفضا برابر 2 و کوچکترین مقدار x_1 برای این چهار زیرفضا برابر -3 است از این رو کران‌های x_1 را به صورت $-3 \le x_1 \le 2$ در نظر می‌گیریم.

جدول ۲: خروجی طبقه‌بندی

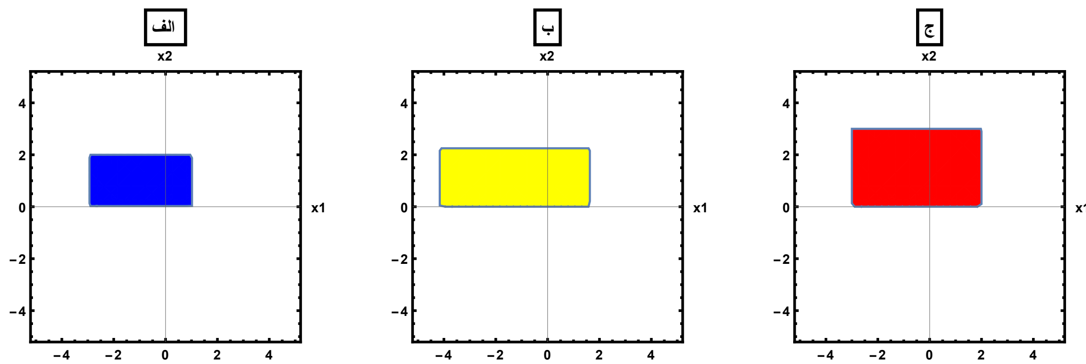
$x_2 < 4.03$
| $x_1 < 1.63$
| | $x_1 < -4.16$: C (4/0)
| | $x_1 \geq -4.16$
| | | $x_2 \leq -4.31$
| | | | $x_1 < -2.62$: B (1/0)
| | | | $x_1 \geq -2.62$: C (4/0)
| | | $x_2 \geq -4.31$
| | | | $x_2 < 0$
| | | | | $x_1 < -3.58$
| | | | | | $x_1 < -4.02$: B (1/0)
| | | | | | $x_1 \geq -4.02$: C (1/0)
| | | | | $x_1 \geq -3.58$: B (6/0)
| | | | $x_2 \geq 0$
| | | | | $x_2 < 2.25$
| | | | | | $x_1 < 0.83$
| | | | | | | $x_1 < -2.04$: A (1/0)
| | | | | | | $x_1 \geq -2.04$
| | | | | | | | $x_1 < -1.13$: B (1/0)
| | | | | | | | $x_1 \geq -1.13$
| | | | | | | | | $x_1 < -0.2$: A (1/0)
| | | | | | | | | $x_1 \geq -0.2$: B (1/0)
| | | | | | $x_1 \geq 0.83$: A (2/0)
| | | | | $x_2 \geq 2.25$
| | | | | | $x_2 < 3.56$
| | | | | | | $x_1 < -0.94$: C (1/0)
| | | | | | | $x_1 \geq -0.94$
| | | | | | | | $x_1 < 0.6$: B (1/0)
| | | | | | | | $x_1 \geq 0.6$: C (1/0)
| | | | | | $x_2 \geq 3.56$: B (2/0)
| $x_1 \geq 1.63$
| | $x_1 < 4.32$: B (11/0)
| | $x_1 \geq 4.32$
| | | $x_2 < -1.68$: D (2/0)
| | | $x_2 \geq -1.68$
| | | | $x_1 < 4.71$: B (1/0)
| | | | $x_1 \geq 4.71$: C (1/0)
$x_2 \geq 4.03$
| $x_1 < -2.5$: D (2/0)
| $x_1 \geq -2.5$: C (3/0)

جدول ۳: خروجی انجمنی

شماره قانون یافت شده	حدود x_1	حدود x_2	کلاس	تعداد جواب‌های صادق در قانون	کسر اطمینان
1	$[4,5]$	$[-4,-3]$	$F=D$	2	conf:(1)
2	$[3,4]$	$[-2,-1]$	$F=B$	2	conf:(1)
3	$[-1,0]$	$[-5,-4]$	$F=C$	2	conf:(1)
4	$[-3,-2]$	$[-5,-4]$	$F=C$	2	conf:(1)
5	$[-4,-3]$	$[4,5]$	$F=D$	2	conf:(1)
6	$[-5,-4]$	$[2,3]$	$F=C$	2	conf:(1)
7	$[4,5]$	$[-2,-1]$	$F=B$	1	conf:(1)
8	$[4,5]$	$[2,3]$	$F=B$	1	conf:(1)
9	$[4,5]$	$[1,2]$	$F=B$	1	conf:(1)
10	$[4,5]$	$[-1,0]$	$F=C$	1	conf:(1)
11	$[3,4]$	$[-4,-3]$	$F=B$	1	conf:(1)
12	$[3,4]$	$[-1,0]$	$F=B$	1	conf:(1)
13	$[3,4]$	$[4,5]$	$F=C$	1	conf:(1)
14	$[-1,0]$	$[2,3]$	$F=B$	1	conf:(1)
15	$[-3,-2]$	$[-3,-2]$	$F=B$	1	conf:(1)
16	$[-3,-2]$	$[1,2]$	$F=A$	1	conf:(1)
17	$[-4,-3]$	$[-4,-3]$	$F=B$	1	conf:(1)
18	$[-4,-3]$	$[-5,-4]$	$F=B$	1	conf:(1)
19	$[-4,-3]$	$[-1,0]$	$F=B$	1	conf:(1)
20	$[-4,-3]$	$[-2,-1]$	$F=C$	1	conf:(1)
21	$[-5,-4]$	$[-2,-1]$	$F=B$	1	conf:(1)
22	$[-5,-4]$	$[0,1]$	$F=C$	1	conf:(1)
23	$[-5,-4]$	$[-3,-2]$	$F=C$	1	conf:(1)
24	$[2,3]$	$[-1,0]$	$F=B$	1	conf:(1)
25	$[2,3]$	$[0,1]$	$F=B$	1	conf:(1)
26	$[2,3]$	$[-3,-2]$	$F=B$	1	conf:(1)
27	$[1,2]$	$[-5,-4]$	$F=B$	1	conf:(1)
28	$[1,2]$	$[-1,0]$	$F=B$	1	conf:(1)
29	$[1,2]$	$[-3,-2]$	$F=B$	1	conf:(1)
30	$[1,2]$	$[3,4]$	$F=B$	1	conf:(1)
31	$[1,2]$	$[2,3]$	$F=C$	1	conf:(1)
32	$[1,2]$	$[0,1]$	$F=A$	1	conf:(1)
33	$[0,1]$	$[1,2]$	$F=B$	1	conf:(1)
34	$[0,1]$	$[3,4]$	$F=B$	1	conf:(1)
35	$[0,1]$	$[4,5]$	$F=C$	1	conf:(1)
36	$[0,1]$	$[2,3]$	$F=A$	1	conf:(1)
37	$[-2,-1]$	$[-5,-4]$	$F=B$	1	conf:(1)
38	$[-2,-1]$	$[-1,0]$	$F=B$	1	conf:(1)
39	$[-2,-1]$	$[0,1]$	$F=B$	1	conf:(1)
40	$[-2,-1]$	$[4,5]$	$F=C$	1	conf:(1)
41	$[-2,-1]$	$[3,4]$	$F=C$	1	conf:(1)
42	$[-2,-1]$	$[1,2]$	$F=A$	1	conf:(1)

به‌طور مشابه برای x_2 داریم: $0 \leq x_2 \leq 3$. اکنون با این رویکرد، فضای حاصل فقط حدود 15 درصد از کل فضای جستجو است.

سه فضای جستجوی کاهش‌یافته حاصل از سه روش متفاوت داده‌کاوی در شکل ۳ نمایش داده شده است. فضاهای کاهش‌یافته حاصل از دو روش طبقه‌بندی و انجمنی، دقیقا شامل جواب بهینه سراسری مسئله است. اکنون به نظر می‌رسد که احتمال یافتن جواب بهینه سراسری در این فضاهای جستجوی کوچکتر بسیار بیشتر از احتمال یافتن آن در کل فضای جستجو باشد.



شکل ۳: مقایسه سه فضای جستجوی کاهش‌یافته حاصل از سه روش الف) خوشه‌بندی، ب) طبقه‌بندی و ج) انجمنی

۵. الگوریتم ژنتیک و روش‌های داده‌کاوی

دوباره مسئله بهینه‌سازی غیرخطی (۱.۴) را در نظر بگیرید. برای بررسی اثر کاهش فضای جستجو روی نتایج حاصل از الگوریتم ژنتیک، این الگوریتم را دوازده مرتبه روی کل فضای شدنی اولیه و همچنین دوازده مرتبه روی فضاهای جستجوی کاهش‌یافته حاصل از سه روش داده‌کاوی اجرا می‌کنیم. میانگین x_1، x_2 و مقادیر تابع هدف به دست آمده از دوازده مرتبه اجرای الگوریتم ژنتیک، در جدول ۴ ارائه شده است. آشکار

جدول ۴: مقایسه میانگین نتایج به دست آمده در فضاهای جستجوی مختلف

	دقیق	خوشه‌بندی	طبقه‌بندی	انجمنی	فضای شدنی اولیه
x_1	0	−0.004	0.008	0.002	−0.092
x_2	0	0.034	0.030	0.073	0.088
f	0	0.261	0.303	0.364	0.482

است که میانگین مقادیر بهینه تابع هدف در کل فضای شدنی اولیه بدتر از همین مقادیر در فضاهای کاهش‌یافته

است. همچنین با کوچکتر شدن فضاهای جستجو، نتایج دقیق‌تری نسبت به کل فضای شدنی حاصل می‌شود. این مقایسه مزیت‌های استفاده از داده‌کاوی برای کاهش فضای جستجو را نمایان می‌کند.

۶. نتیجه گیری

در این تحقیق، برای دستیابی به هدف کاهش فضای جستجوی یک مسئله بهینه‌سازی غیرخطی و افزایش احتمال یافتن جواب بهینه سراسری، روش‌های خوشه‌بندی، طبقه‌بندی و انجمنی در داده‌کاوی مورد استفاده قرار گرفت. نتایج نشان داد که مساحت فضاهای کاهش‌یافته حاصل، حداکثر 15 درصد از مساحت کل فضای شدنی مسئله است. سپس با استفاده از الگوریتم ژنتیک، مسئله غیرخطی روی کل فضای شدنی اولیه و فضاهای کاهش‌یافته حل شد و بررسی نتایج مشخص کرد که الگوریتم ژنتیک روی فضاهای کاهش‌یافته می‌تواند جواب‌های دقیق‌تری تولید کند. همچنین فضای حاصل از روش خوشه‌بندی کوچکترین فضا است و ترکیب این روش با الگوریتم ژنتیک، بهترین تقریب را برای جواب بهینه سراسری مسئله نسبت به دو فضای کاهش‌یافته دیگر ارائه می‌دهد.[۲۳]

مراجع

[۱] ن. حاجی‌حیدری و ب. خاکباز، *داده‌کاوی*، انتشارات دانشگاه تهران، چاپ دوم، ۱۳۹۲.

[۲] م. غضنفری، س. علیزاده و ب. تیمورپور، *داده‌کاوی و کشف دانش*، انتشارات دانشگاه علم‌وصنعت ایران، چاپ پنجم، ۱۳۹۵.

[۳] م. کانتاردزیک و ا. علیخانزاده، *داده‌کاوی*، انتشارات علوم رایانه، چاپ سوم، ۱۳۹۲.

[۴] ن. مهدوی‌امیری و م. پورکاظمی، *برنامه‌ریزی خطی و غیرخطی*، انتشارات دانشگاه صنعتی شریف، چاپ سوم، ۱۳۹۱.

[5] M.S. Bazaraa and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms*, John Wiley & Sons, 2nd. ed. New York, 1979.

[6] J. Branke and C. Schmidt, *Faster convergence by means of fitness estimation*, Soft. Comput. **9** (2005) 13–20.

[7] T.Y. Chen and J.H. Huang, *Application of data mining in a global optimization algorithm*, Adv. Eng. Software **66** (2013) 1–10.

[8] J. Nocedal and S. Wright, *Numerical Optimization*, Springer-Verlag, 2nd. ed. New York, 2006.

[9] J. Pearl, *Heuristics: Intelligent Search Strategies for Computer Problem Solving*, Addison-Wesley, Boston, MA, 1984.

[10] F.G. Teofilo, *Clustering to minimize the maximum intercluster distance*, Theor. Comput. Sci. **38** (1985) 293–306.

[11] I.H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, San Francisco, CA, 2005.

[12] Q. Zhang, J. Sun J, E. Tsang and J. Ford, *Hybrid estimation of distribution algorithm for global optimization*, Eng. Comput. **21** (2004), no. 1, 91–107.

[۲۳] این مقاله از پایان‌نامه کارشناسی ارشد نویسنده اول استخراج شده است.